AD-A097 541 CLEMSON UNIV SC DEPT OF MATHEMATICAL SCIENCES F/G 12/1
THE ACCURACY OF A MODIFIED PEIZER APPROXIMATION TO THE HYPERGEO--ETC(U)
JUL 80 R F LING, J W PRATT N00014-75-C-0451
UNCLASSIFIED N126 NL

END
DATE
FILMED
DTIC

AD A097541

DEPARTMENT
OF
MATHEMATICAL
SCIENCES

LEVEL II (6)

(6) THE ACCURACY OF A MODIFIED PEIZER APPROXIMATION TO THE HYPERGEOMETRIC DISTRIBUTION, WITH COMPARISONS TO SOME OTHER APPROXIMATIONS.

(10) Robert F. Ling [redacted] John W. Pratt
Clemson University and Harvard University

Department of Mathematical Sciences
Clemson University

(12) 35

(9) Technical Rep[redacted]t #348

(11) Jul[redacted]980

(14) N126, TR-348

This work was supported in part by the Office of Naval Research under Contract N00014-75-C-0451 (15)

ABSTRACT

Results of an extensive empirical study of the accuracy of seven normal and three binomial approximations to the hypergeometric distribution are presented in terms of maximum absolute error under various conditions on the variables. The most useful condition are provided by the minimum cell in the given or complementary 2×2 table and the tail probability itself. Of the normal approximations, a modification on one due to Peizer is far the best. It has error at most .0001, for example, if the minimum cell is at least 9, or if the tail probability is below .01 and the minimum cell is at least 4. Especially detailed results are given for this approximation.

Key words: maximum absolute error, hypergeometric distribution, normal approximation.

AUTHORS' FOOTNOTE

Robert F. Ling is Professor, Department of Mathematical Sciences, Clemson University, Clemson, SC 29631. John W. Pratt is Professor, Graduate School of Business Administration, Harvard University, Boston, MA 02163. This work was supported in part by the Office of Naval Research under Contract N00014-75-C-0451 and by the Associates of the Harvard Business School.

1. Introduction

This paper reports results from an empirical study of several normal and binomial approximations to the hypergeometric distribution. The motivation for considering approximations is that machine computation of an exact formula is often inefficient, because of the number of terms required, and is sometimes infeasible because of overflow or underflow in machine arithmetic. Furthermore, even tables as large as Lieberman and Owen (1961) are inevitably inconvenient and incomplete, and they cannot be made part of statistical computing packages. An empirical study is needed because exact results on the accuracy of most approximations are intractible to obtain theoretically and the empircal knowledge available is very limited. Indeed, it is nonexistent for the best normal approximation studied here.

The performance criterion is essentially maximum absolute error under certain conditions on the variables. Advantages of absolute over relative error are that it is more often wanted in practical problems and that it enables one to guarantee the numerical accuracy of the approximated probabilities to a specified precision, such as k decimal places, as in Ling (1978). As a refinement, we considered the maximum absolute error in several ranges of the tail probability. This permits one to get a feel for other criteria, such as relative error, also.

Five normal approximations were investigated: the usual $\frac{1}{2}$-corrected chi statistic, three other normal approximations studied by Molenaar (1970), and a modification of an approximation due to Peizer (1966?; see Section 5). Binomial approximations are not appropriate competitors to normal approximations, since binomial tails present almost the same computational problems as hypergeometric tails, merely reducing the number of variables from four to three. For interest, however, we investigated Wise's (1954) one-term binomial approximation and two refinements studied by Molenaar (1970).

The notation and approximations are defined in Section 2. Some comparisons are given in Section 3. Because the modified Peizer approximation is both far superior to the other normal approximations and simple to compute, considerable additional information on its accuracy is provided in Section 4. This information took at least 15 hours of CPU on an IBM 3033 computer and hence the expense of obtaining comparably detailed information for other approximations would not be justified. Section 5 gives the rationale in Peizer's approximation and its modification. Section 6 contains information about our calculation and search procedures.

2. Notation and Approximations

Given the 2 2 Table with fixed margins:

$$\begin{array}{cc|c} a & b & n \\ c & d & m \\ \hline r & s & N \end{array} \quad , \; n+m = r+s = N,$$

the associated hypergeometric cumulative probability is

$$P(X \leq a|n,r,N) = \sum_{j=0}^{a} \binom{n}{j}\binom{m}{r-j} \Big/ \binom{N}{r}. \tag{2.1}$$

We consider approximations $\Phi(z)$ where Φ is the unit normal cumulative and z is one of the following approximate normal deviates. The first is the square-root of the usual $\frac{1}{2}$ - corrected chi-square statistic,

$$\chi = (a + \tfrac{1}{2} - nr/N)/(mnrs/N^3)^{1/2} . \tag{2.2}$$

Substituting the exact standard deviation in the denominator gives

$$u = (a + \tfrac{1}{2} - nr/N)N/(mnrs/(N-1))^{1/2}. \tag{2.3}$$

Molenaar (1970, p. 120, equation 2.5) expands the exact normal deviate to third order as

$$\begin{aligned} z_1 = \chi &+ (m-n)(s-r)(1-\chi^2)/6(mnrs/N)^{1/2} \\ &+ \{\chi^3(5N^2-14mn-14rs+38mnrs/N^2) \\ &+ \chi(-2N^2+2mn+2rs+10mnrs/N^2)\}N/72mnrs. \end{aligned} \tag{2.4}$$

Molenaar also develops and investigates square root approximations, recommending (p. 133)

$$z_2 = 2((a+1)^{1/2}(d+1)^{1/2} - b^{1/2}c^{1/2})/(N-1)^{1/2} \tag{2.5}$$

near the customary significance levels and

$$z_3 = 2((a + \tfrac{3}{4})^{1/2}(d + \tfrac{3}{4})^{1/2} - (b - \tfrac{1}{4})^{1/2}(c - \tfrac{1}{4})^{1/2}/N^{1/2} \tag{2.6}$$

in the middle of the distribution. He also investigates adjusting χ by variable continuity corrections and added correction terms, obtaining as his most accurate recommended approximate normal deviate (p. 136)

$$\begin{aligned} z_4 = \chi &+ (1-\chi^2)[(m-n)(s-r)/6(mnrs/N)^{1/2} \\ &- \chi(N^2-3mn)N/48mnrs] . \end{aligned} \tag{2.7}$$

A modification of an approximation due to Peizer (see Section 5) is

$$z_5 = \frac{a'd'-b'c'}{|AD-BC|}\left(\frac{2mnrsN'}{m'n'r's'N}L\right)^{1/2} \tag{2.8}$$

where A = a+.5, B = b-.5, C = c-.5, and D = d+.5 are the $\frac{1}{2}$- corrected entries,

$$a' = A + \frac{1}{6} + \frac{.02}{A+.5} + \frac{.01}{n+1} + \frac{.01}{r+1}, \tag{2.9}$$

and similarly for b', c', and d' with n and r replaced by the row and column total for the entry in question, $m' = m + \frac{1}{6}$, $n' = n + \frac{1}{6}$, $r' = r + \frac{1}{6}$, $s' = s + \frac{1}{6}$, $N' = N - \frac{1}{6}$, and

$$L = A \log \frac{AN}{nr} + B \log \frac{BN}{ns} + C \log \frac{CN}{mr} + D \log \frac{DN}{ms}, \tag{2.10}$$

all logarithms being natural, and their arguments being the (corrected) observed over "expected" cell frequencies. The modified Peizer approximation can also be expressed in terms of the function g defined and tabulated in Peizer and Pratt (1968) as

$$z_5 = (a'd'-b'c')(N'G/m'n'r's')^{1/2} \tag{2.11}$$

where

$$G = 1+\{ms\cdot g(\frac{AN}{nr}) + mr\cdot g(\frac{BN}{ns}) + ns\cdot g(\frac{CN}{mr}) + nr\cdot g(\frac{DN}{ms})\}/N^2; \tag{2.12}$$

$$g(x) = (1-x^2+2x \log x)/(1-x)^2, \quad 0 < x \neq 1,$$

$$g(0) = 1,\ g(1) = 0. \tag{2.13}$$

Noting the probability of b or more is 1 minus the probability of b-1 or less and exchanging columns shows that

$$P(X \leq a|n,r,N) = 1-P(X \leq b|n,s,N). \tag{2.14}$$

Since all the normal approximations above transform appropriately under such an exchange of columns, or a similar exchange of rows, or an exchange of rows with columns, we can without loss of generality arrange the table so that

$$a \leq d \quad \text{and} \quad a < b \leq c \tag{2.15}$$

or equivalently

$$2a+1 \leq n \leq r \leq N-n. \tag{2.16}$$

To present our results, we therefore introduce

$$k = \min(a,b-1,c-1,d), \tag{2.17}$$

and we let n and r denote the associated margins, with $n \leq r$. Then (2.16) holds with $a = k$.

Very small values of k are of little interest in comparing approximations because the exact probability is easily calculated as a sum of k+1 terms by (2.1), which may be rewritten as

$$P(x \leq k|n,r,N) = [1 + \frac{rn}{m-r+1} + \ldots$$

$$+ \frac{r(r+1) \quad \ldots \quad (r+k-1)}{(m-r+1) \quad \ldots \quad (m-r+k)}\binom{n}{k}]\binom{m}{r}/\binom{N}{n}. \tag{2.18}$$

Binomial approximations belong in a different category from normal approximations, for several reasons. Binomial tables are far bulkier and less complete than normal tables. For machine work, hypergeometric tails are often as easy to compute directly as binomial tails. When an approximation to the hypergeometric distribution is needed, a binomial approximation would itself usually need to be approximated. The modified Peizer approximation to the hypergeometric distribution is already an adaptation of the refined Peizer-Pratt normal approximation to the binomial, which Ling (1978) found substantially better than others. Binomial approximations therefore cannot appropriately be regarded as competitors to normal approximations. We nevertheless considered three binomial approximations. All are to be applied after rearrangement of the 2×2 table so that n is the smallest margin ($n \leq r \leq N-n$) but transform appropriately when columns are exchanged (so that the remaining inequality in (2.16) is no loss of generality). All approximate the hypergeometric tail by a binomial tail with the same n and the same number of occurrences a but with p depending on a as well as on the margins. The first is the first term of Wise's (1954) series and takes for p

$$p_1 = \frac{2r-a}{2N-n+1} . \tag{2.19}$$

The second is a modification of this developed by Molenaar (1970) as an approximation to Wise's (1954) second-order approximation and uses

$$p_2 = p_1 + [(n+1)(ap_1-(b-1)(1-p_1)) - a(a+2)p_1^{-1} + (b^2-1)(1-p_1)^{-1}]/6(2N-n+1)^2. \tag{2.20}$$

The third is an alternative proposed by Molenaar as simpler than p_2 but usually close to it and almost as good:

$$p_3 = p_1+2n(rn/N-a-\tfrac{1}{2})/3(2N-n+1)^2. \tag{2.21}$$

Molenaar finds other binomial approximations inferior to these.

3. Comparison of Approximations

We first compared the maximum absolute error of each of the approximations defined above, as a function of N, in the region ($1 \leq n \leq r \leq N/2$, $0 \leq a < n$) corresponding to the entries tabulated by Lieberman and Owen (1961, pp. 33-293). Table 1 gives the maximum error for selected N (≤ 50) and the error graphs of six of the approximations are given in Figure A. The maxima decrease slowly, if at all, as a function of N, with poor error bounds. Examination of the detailed results revealed that all of the maxima for the normal approximations occur at $a = 0$ (hence $k = 0$), a case of almost no interest; while the maximum of the binomial approximations occur at nonzero values of k. We conclude that it is far more useful to fix k than N in tabling maximum errors and comparing approximations.

Table 2 gives, for the same approximations, the maximum error which can occur for $k = 4$ and 8 in two ranges of N. For $k = 4$, $N \leq 200$, for instance, this is the maximum error for all 2×2 tables with $\min(a,b-1,c-1,d) = 4$ and $a+b+c+d \leq 200$. The columns of Table 2 give results for restricted ranges of the smaller tail probability P. The dependence of the error on other variables such as r and n, is complicated and different for different approximations, so we have not attempted to present it. Our main conclusions from Table 2 are:

1. The modified Peizer approximation is more accurate than all other normal approximations by at least an order of magnitude and is by far the best bet for any ordinary machine calculation.
2. The results in this Table, together with various other schemes of tabulation we have explored (by fixing various combinations of (k,n,r,N)), suggest that the most important variables are k and the tail probability.

3. For k fixed, the normal approximations, with the exception of Molenaar adjusted χ (2.7) and modified Peizer (2.8, 2.11), have increasing maximum error as N increases. The binomial approximations generally perform well when N is large, even when k is relatively small.
4. The adjustment of the denominator between χ and u is insignificant.
5. Molenaar's finding that adjusting the square root (2.6) helps in the middle of the distribution is only partially borne out.
6. Molenaar's adjustment of χ (2.7) is superior to use of the expansion (2.4) he gives (which he does not propose as an approximation).
7. The best of the binomial approximations is Molenaar's approximation (2.20) to the second-order Wise approximation. It is inferior to the modified Peizer approximation in the smaller range of N but superior in the larger range (where $N > 50k$).

Binomial approximations should, of course, work well when N is large compared to n, that is, the sampling fraction is small, and they become exact as $N \to \infty$ if $n/N \to 0$.

4. Accuracy of the Modified Peizer Approximation

Since the modified Peizer approximation is clearly superior to the others we looked at for most purposes, we explored its accuracy considerably further. Full presentation of a complicated function of four variables being impossible, we give results in several forms.

Table 3 and Figure B extend Table 2 to the range $0 \leq k \leq 50$, giving the maximum absolute error of the modified Peizer approximation for k fixed, with no other restriction except on the tail probability. In particular, the absolute error is less than .0001 for all combinations of variables (all 2×2 tables) with $k \geq 9$; for tail probabilities less than .01, $k \geq 4$ suffices. Table 4 shows such values of k for various standards of accuracy.

The values in Table 3 for $k \geq 4$ can be fitted extremely closely by choosing an appropriate linear function of k, log k, and $(\log k)^2$ for each range of tail probabilities separately. The coefficients of such functions obtained by unweighted regression of the values shown are given at the foot of the table. All values fit within .08% except for tail probabilities between .01 and .1, where the fit is within .7% (.2% for $k \geq 24$). Since direct calculation is easy and the approximation poorer for $k \leq 3$, these values were excluded from the fit.

Constraining other variables in addition to k of course reduces the maximum possible error. As an illustration, we exhibit the maximum as a function of (n,N) for k = 8 in Table 5 and a corresponding contour plot in Figure C. The pattern for other values of k is similar.

As another illustration, Figure D shows the behavior of the error as a function of r and n for N = 50 by means of error contours. Since most of the contours never reach the axes, it appears difficult to find restrictions on r and n which would bound the error usefully. Moreover, error bounds based on r and/or n would probably be unacceptably large because most of the maxima occur at small values of k.

5. Origin and Rationale for Peizer's Approximation and Its Modification

David Peizer (1966?) in handwritten notes[1], extends his joint work with Pratt (1968) to the hypergeometric distribution as follows. He arrives, apparently by a combination of analysis, analogy, and inspiration, at an approximate normal deviate of the form (2.11) with $a' = A+c_1$, $m' = m+c_2$, and similarly for the other entries and margins, and $N' = N+c_3$. By asymptotic expansion near the median, he finds that the best constants are $c_1 = c_2 = \frac{1}{6}$, $c_3 = -\frac{1}{6}$. To express (2.11) in terms of logs, one can use

$$1+qg(P/p) + pg(Q/q) = 2pq(P \log(P/p) + Q \log (Q/q))/(P-p)^2 \qquad (5.1)$$

which holds for all nonnegative p, q, P, and Q with $p+q = P + Q = 1$ and can be obtained from (1.2) of Peizer and Pratt (1968) or otherwise. Applying (5.1) once with $p = r/N$, $P = A/n$ and once with $p = r/N$, $P = C/m$ gives, after some algebra,

$$G = 2mnrs\ L/N(AD-BC)^2 \qquad (5.2)$$

where L is given by (2.10). Substituting (5.2) in (2.11) gives (2.8).

In the binomial limiting case, Peizer's approximation reduces to the simpler of Peizer and Pratt's (1968) approximations. It can be modified so as to reduce to their refined approximation in many ways, of which the simplest is to add $.01/(n+1) + .01/(r+1)$ to a' and similarly for b', c', and d'. This is what we did in (2.9).

Calculation with the resulting approximation indicates that, at the maximum absolute error over all probability classes, the tail probabilities are usually too small and that adjustment of order $N^{-1.5}$ might help. Adding a constant to N' does not give an adjustment of this order, but adding the same multiple of 1/N to a', b', c', and d' does, because of cancellation. Of course, any term of order 1/N vanishes in the binomial limiting case. The choice -.08/N fared well in a few cases we looked at, reducing the maximum absolute error by more than 30% in the central proability classes with k fixed, but at the expense of an increase in the extreme probability classes. We did not investigate further refinement along these lines at all extensively, but it might be useful under some circumstances, especially when the main focus of attention is maximum absolute error over all probability classes.

One other modification we tried was to replace the numerators .02 and .01 in (2.9) by the values which minimize the maximum asymptotic absolute error in the binomial case. These values are $(16+v)/810 = .0200969$ and $(8+23v)/810 = .0177836$, where $v = .278465$ is the solution of $ev = e^{-v}$. Replacing .02 and .01 by these values reduces the maximum asymptotic absolute error by about 22% for all binomial and Poisson distributions. It made no appreciable difference, however, in the nonasymptotic, hypergeometric computer runs we carried out, and we gave it up. The asymptotically minimax values can be derived by observing that the asymptotic absolute error is of the form $C_1|z^2-C_2|e^{-z^2/2}$, by Pratt (1968 (5.10)), and that the maximum of this with respect to z is minimized by $C_2 = 2v$. Calculations like those of Pratt (1968, Section 5.2) show that $C_2 = 2v$ for the values given above, and the reduction achieved is derived by further, similar calculation.

6. Computational Considerations

6.1 Computational Precision and Machine Dependence

All numerical results reported in this study were computed in double-precision on an IBM 3033 machine using FORTRAN programs compiled by the extended-H compiler (with level 2 optimization). Since machine-dependent roundoff errors occur at decimal digits well beyond those reported, the approximation errors can be attributed entirely to the quality of the approximation formulas. Thus, the results are machine-dependent only to the extent of possible dependence on the word lengths and floating point softwares of various machines. The reported results may not hold for computations performed in single-precision arithmetic or on machines with word lengths considerably shorter than what was actually used.

6.2 Computation of "Exact" Hypergeometric Probabilities

Let $p(x) = p(x,n,r,N)$ denote the hypergeometric point probabilities in (2.1), i.e.,

$$p(x) = p(x,n,r,N) = \binom{n}{x}\binom{N-n}{r-x}/\binom{N}{r}$$

$$= \frac{n!r!}{(n-x)!(r-x)!x!} \frac{(N-n)!(N-r)!}{N!(N-n-r+x)!}, \quad 0 \le x \le n. \tag{6.1}$$

Direct computations of these probabilities on the computer are frequently infeasible either because of "overflow" in the computation of factorials or "underflow" in the resulting $p(x)$. For example, $p(100, 500, 500, 1000)$ requires the computation of $(500!)^4/((400!)^2(100!)^2(1000!))$. The smallest of these factorials, 100!, is of the order 10^{157} which greatly exceeds the maximum number directly computable on the IBM 3033 machine (about 10^{76}) or on most 32-bit word-length machines; while the probability $p(100)$ is of the order 10^{-84} which if computed by other methods would cause "underflow" for being excessively small.

Lieberman and Owen (1961) calculate their tabled values by making use of a computer stored table of Log N! for N = 1(1)2000, with 15 digits in the mantissa. Presumably they did not convert $\log(p(x))$ to $p(x)$ when the log is a large negative number. Although they claim their computed probability results to be accurate to at least eight decimal places (Lieberman and Owen 1961, p. 4), with six decimal places tabulated, we found (on checking only the case N = 20) 9 erroneous entries in the cumulative probabilities for (x,n,r) = (4,5,5), (4,5,7), (5,6,8), (5,8,9), (7,8,9), (4,9,9), (3,6,10), and (9,10,10). In each case, the last digit of the tabled value is less than the correct value by 1.

We calculated our "exact" probabilities from (6.1), by the recursion

$$p(x+1,n,r,N) = p(x,n,r,N)\,\frac{(n-x)(r-x)}{(x+1)(N-n-r+x+1)} \quad \text{for } x \geq 0, \tag{6.2}$$

where

$$p(0,n,r,N) = \frac{(N-n)(N-n-1)\ldots(N-n-r+1)}{N(N-1)\ldots(N-r+1)} .$$

A special FORTRAN subroutine was written for the calculation of (6.2) so that double-precision (about 15 significant digits) is maintained regardless of the magnitude of the point probabilities. Thus, even if p(x) is of the order $10^{-1,000,000}$, it is computed, although we do not cumulate the point probabilities in (2.1) for $p(x) < 10^{-15}$. We are reasonably sure that all the numerical results in this article are correct in all the digits reported since computations were performed in double-precision and the smallest error reported is of the order 10^{-10}.

6.3 Search for Maximum Errors

The searches made for the comparison of the approximations in Tables 1 and 2 were exhaustive.

For further exploration of the modified Peizer approximation an exhaustive search was first made as far as the values of N shown in Table 6. For small values of k, examination of the detailed output strongly indicated that the maximum error had long since been passed in each interval of tail probabilities. Furthermore, the value of N/k at the maximum tended to decrease with k in each interval and was less than 28 for $8 \leq k \leq 16$. (See Table 7 for $k = 8$ and 16.) Also r-n at the maximum never exceeded 4 for $k \leq 16$ and never exceeded 6 in any situation where the maximum appeared to have been reached, except that, for tail probabilities between .01 and .05, the maximum for $k \leq 28$ occurred at $r = N$-n, r-$n \leq 8$, $N \leq 7k$. Accordingly, for each $k \geq 18$, the search was extended at least as far as $N = 30k$ but with the added restriction r-$n \leq 12$, the computer time for exhaustive search being prohibitive for large values of N. The maxima found thereby for $k \geq 18$ all occurred at $N < 27.5k$ and r-$n \leq 7$. The only surprise was that, for tail probabilities between .01 and .05, the maximum switched from one tail to the other between $k = 28$ and $k = 32$, while the maximizing N switched simultaneously from about $4.8k$ to about $27.4k$ with no change in the pattern or r-n but now $r \neq N$-n.

The numerical evidence convinces us that the search was adequate. It is not surprising that the maximum should occur near $r = n$, for the simple reason that this is one of the two extreme types of 2×2 table possible. Furthermore, the other extreme is the binomial limit, where the modified Peizer approximation reduces to the refined Peizer-Pratt approximation. The accuracy at the binomial limit is better than at $r = n$ by a factor of 3 or so. Presumably r is not exactly n at the maximum because of discreteness. Specifically, there is a trade-off between coming close to the extreme $r = n$ and coming close to the valeu of r/N which would be worst in a continuous version of the problem.

REFERENCES

Lieberman, Gerald J., and Owen, Donald B. (1961), "Tables of the Hypergeometric Probability Distribution", Applied Mathematics and Statistics Laboratories Technical Report No. 50, Stanford University.

Ling, Robert F. (1978), "A Study of the Accuracy of Some Approximations for t, χ^2, and F Tail Probability", Journal of the American Statistical Association, 73, 274-283.

Molenaar, Wouter. (1970), Approximations to the Poisson, Binomial and Hypergeometric Distribution Functions, Mathematical Centre Tracts 31. Amsterdam.

Peizer, David B. (1966?), unpublished manuscript, Harvard University.

Peizer, David B., and Pratt, John W. (1968), "A Normal Approximation for Binomial, F, Beta, and Other Common Related Tail Probabilities, I," Journal of the American Statistical Association, 63, 1416-1456.

Pratt, John W. (1968), "A Normal Approximation for Binomial, F, Beta, and Other Common, Related Tail Probabilities, II," Journal of the American Statistical Association, 63, 1457-1483.

Wise, M. E. (1954), "A Quick Convergent Expansion for Cumulative Hypergeometric Probabilities, Direct and Inverse," Biometrika, 41, 317-329.

FOOTNOTE

1 These notes are in Pratt's possession and almost surely precede August 1966, when Peizer left Harvard [?]. We have been unable to locate him. He submitted a paper to JASA in March, 1968. It was returned for revision but never resubmitted. Pratt has the correspondence but no copy of the paper.

TABLE 1

Maximum Absolute Error (x 10,000) of Approximations to the Hypergeometric Distribution for Fixed N at Selected Values

	N =	5	10	15	20	25	30	35	40	45	50
Normal Approximations											
1/2-corrected x	(2.2)	119	439	533	616	650	694	724	748	765	780
1/2-corrected u	(2.3)	266	479	558	624	671	705	728	757	769	787
Molenaar expansion	(2.4)	191	914	667	847	796	820	833	837	845	798
Molenaar square root	(2.5)	413	444	510	594	663	707	747	788	814	837
Molenaar alt. sq. root	(2.6)	559	259	352	399	425	442	453	461	467	471
Molenaar adjusted x	(2.7)	134	181	487	495	400	379	485	488	444	438
Modified Peizer	(2.8, 2.11)	82	50	48	47	46	45	45	45	44	44
Binomial Approximations											
Wise	(2.19)	111	159	131	144	135	136	134	148	140	147
Molenaar modified Wise	(2.20)	4	15	11	14	11	12	11	13	12	13
Molenaar alt. mod. Wise	(2.21)	36	40	26	27	20	21	17	20	17	18

TABLE 2

Maximum Absolute Error (x 100,000) of Approximations to the Hypergeometric Distribution for Fixed k at k = 4 and k = 8

k = 4		min(P,1-P) ≤	.5	.10	.05	.010	.005	.0010	.0005	.0001
		>	.1	.05	.01	.005	.001	.0005	.0001	0
1/2-corrected		N≤200	1874	957	984	595	370	94	55	17
x	(2.2)	200<N≤500	2348	1134	1158	674	408	141	90	30
1/2-corrected		N≤200	1882	919	950	586	367	94	58	18
u	(2.3)	200<N≤500	2350	1119	1145	671	407	143	91	30
Molenaar		N≤200	339	681	879	729	463	100	50	10
expansion	(2.4)	200<N≤500	327	700	940	779	480	100	50	10
Molenaar		N≤200	3361	920	364	274	245	138	102	48
square root	(2.5)	200<N≤500	4272	1262	505	310	279	159	118	57
Molenaar		N≤200	806	734	743	547	422	205	147	67
alt. sq. root	(2.6)	200<N≤500	1034	846	855	638	491	241	174	82
Molenaar		N≤200	169	272	289	215	166	70	43	10
adjusted x	(2.7)	200<N≤500	277	351	354	212	102	54	42	10
Modified		N≤200	33	24	16	8	6	3	2	1
Peizer	(2.8, 2.11)	200<N≤500	13	12	11	8	6	3	2	1
Wise		N≤200	1488	1257	974	415	262	80	53	12
binomial	(2.19)	200<N≤500	109	101	84	36	22	7	4	1
Molenaar		N≤200	135	114	85	38	23	7	5	1
modified Wise	(2.20)	200<N≤500	1	1	1	<1	<1	<1	<1	<1
Molenaar		N≤200	251	210	163	73	43	13	9	2
alt.mod. Wise	(2.21)	200<N≤500	30	29	25	11	7	2	1	<1

TABLE 2 (contd.)

Maximum Absolute Error (x 100,000) of Approximations to the Hypergeometric Distribution for Fixed k at k = 4 and k = 8

k = 8		min(P,1-P) ≤	.5	.10	.05	.010	.005	.0010	.0005	.0001
		>	.1	.05	.01	.005	.001	.0005	.0001	0
1/2-corrected		N≤200	1016	535	550	357	242	73	41	9
x	(2.2)	200<N≤500	1460	721	736	462	300	87	55	17
1/2-corrected		N≤200	1020	495	518	345	237	72	40	9
u	(2.3)	200<N≤500	1462	705	723	458	298	89	56	18
Molenaar		N≤200	145	262	315	285	224	83	47	10
expansion	(2.4)	200<N≤500	141	260	328	307	246	89	48	10
Molenaar		N≤200	1902	464	174	128	116	60	42	17
square root	(2.5)	200<N≤500	2762	773	297	162	146	76	54	22
Molenaar		N≤200	461	381	386	273	206	89	60	23
alt. sq. root	(2.6)	200<N≤500	677	487	492	351	261	113	77	30
Molenaar		N≤200	90	106	169	127	92	34	21	6
adjusted x	(2.7)	200<N≤500	100	119	120	82	57	14	8	4
Modified		N≤200	10	8	5	3	2	1	1	<1
Peizer	(2.8, 2.11)	200<N≤500	5	5	4	3	2	1	1	<1
Wise		N≤200	1476	1280	975	397	256	75	49	12
binomial	(2.19)	200<N≤500	199	191	153	63	41	13	8	2
Molenaar		N≤200	128	109	36	33	21	6	4	1
modified Wise	(2.20)	200<N≤500	3	3	2	1	1	<1	<1	<1
Molenaar		N≤200	197	169	134	52	34	10	7	1
alt.mod. Wise	(2.21)	200<N≤500	42	41	33	13	9	3	1	<1

TABLE 3

Maximum Absolute Error of Modified Peizer
Approximation to the Hypergeometric D:stribution

min(P,1-P)	≤ .5 > .1	.10 .05	.05 .01	.010 .005	.005 .001	.0010 .0005	.0005 .0001	.0001 0
k								
0	$.0^{2}92$	$.0^{2}65$	$.0^{2}40$	$.0^{2}26$	$.0^{2}21$	$.0^{3}20$	$.0^{3}14$	$.0^{4}57$
1	$.0^{2}21$	$.0^{2}13$	$.0^{3}90$	$.0^{3}53$	$.0^{3}44$	$.0^{3}21$	$.0^{3}14$	$.0^{4}20$
2	$.0^{3}91$	$.0^{3}61$	$.0^{3}43$	$.0^{3}23$	$.0^{3}18$	$.0^{4}85$	$.0^{4}57$	$.0^{4}20$
3	$.0^{3}50$	$.0^{3}39$	$.0^{3}25$	$.0^{3}13$	$.0^{3}10$	$.0^{4}45$	$.0^{4}30$	$.0^{4}10$
4	$.0^{3}33$	$.0^{3}24$	$.0^{3}16$	$.0^{4}84$	$.0^{4}64$	$.0^{4}28$	$.0^{4}18$	$.0^{5}64$
5	$.0^{3}23$	$.0^{3}19$	$.0^{3}11$	$.0^{4}60$	$.0^{4}45$	$.0^{4}19$	$.0^{4}13$	$.0^{5}42$
6	$.0^{3}17$	$.0^{3}13$	$.0^{4}94$	$.0^{4}45$	$.0^{4}34$	$.0^{4}14$	$.0^{5}92$	$.0^{5}31$
7	$.0^{3}13$	$.0^{3}11$	$.0^{4}70$	$.0^{4}36$	$.0^{4}26$	$.0^{4}11$	$.0^{5}70$	$.0^{5}23$
8	$.0^{3}10$	$.0^{4}82$	$.0^{4}53$	$.0^{4}29$	$.0^{4}21$	$.0^{5}87$	$.0^{5}56$	$.0^{5}18$
9	$.0^{4}84$	$.0^{4}70$	$.0^{4}47$	$.0^{4}24$	$.0^{4}18$	$.0^{5}71$	$.0^{5}45$	$.0^{5}15$
10	$.0^{4}69$	$.0^{4}55$	$.0^{4}37$	$.0^{4}20$	$.0^{4}15$	$.0^{5}59$	$.0^{5}38$	$.0^{5}12$
11	$.0^{4}58$	$.0^{4}48$	$.0^{4}34$	$.0^{4}18$	$.0^{4}13$	$.0^{5}50$	$.0^{5}32$	$.0^{5}10$
12	$.0^{4}50$	$.0^{4}43$	$.0^{4}28$	$.0^{4}15$	$.0^{4}11$	$.0^{5}44$	$.0^{5}28$	$.0^{6}89$
13	$.0^{4}43$	$.0^{4}36$	$.0^{4}25$	$.0^{4}13$	$.0^{5}97$	$.0^{5}38$	$.0^{5}24$	$.0^{6}77$
14	$.0^{4}38$	$.0^{4}32$	$.0^{4}23$	$.0^{4}12$	$.0^{5}86$	$.0^{5}34$	$.0^{5}21$	$.0^{6}68$
15	$.0^{4}33$	$.0^{4}29$	$.0^{4}19$	$.0^{4}11$	$.0^{5}77$	$.0^{5}30$	$.0^{5}19$	$.0^{6}60$
16	$.0^{4}30$	$.0^{4}25$	$.0^{4}18$	$.0^{5}97$	$.0^{5}69$	$.0^{5}27$	$.0^{5}17$	$.0^{6}54$
18	$.0^{4}24$	$.0^{4}21$	$.0^{4}14$	$.0^{5}80$	$.0^{5}57$	$.0^{5}22$	$.0^{5}14$	$.0^{6}44$
20	$.0^{4}20$	$.0^{4}17$	$.0^{4}12$	$.0^{5}68$	$.0^{5}48$	$.0^{5}18$	$.0^{5}12$	$.0^{6}36$
24	$.0^{4}14$	$.0^{4}12$	$.0^{5}87$	$.0^{5}51$	$.0^{5}36$	$.0^{5}14$	$.0^{6}85$	$.0^{6}26$
28	$.0^{4}11$	$.0^{5}91$	$.0^{5}68$	$.0^{5}40$	$.0^{5}28$	$.0^{5}11$	$.0^{6}66$	$.0^{6}20$
32	$.0^{5}82$	$.0^{5}71$	$.0^{5}53$	$.0^{5}33$	$.0^{5}23$	$.0^{6}85$	$.0^{6}53$	$.0^{6}16$
36	$.0^{5}65$	$.0^{5}57$	$.0^{5}45$	$.0^{5}27$	$.0^{5}19$	$.0^{6}70$	$.0^{6}43$	$.0^{6}13$
40	$.0^{5}54$	$.0^{5}46$	$.0^{5}38$	$.0^{5}23$	$.0^{5}16$	$.0^{6}59$	$.0^{6}36$	$.0^{6}11$
50	$.0^{5}35$	$.0^{5}31$	$.0^{5}27$	$.0^{5}16$	$.0^{5}11$	$.0^{6}41$	$.0^{6}25$	$.0^{7}76$

coefficient of curve fitted to log (max absolute error)
in each class separately for $4 \le k \le 50$

k	.00454	.00686	.02391	.00322	.00303	.00331	.00379	.00295
logk	-1.372	-1.148	-.9002	-1.468	-1.552	-1.676	-1.714	-1.834
$(\log k)^2$	-.0959	-.1360	-.2168	-.0296	-.0201	-.0107	-.0099	.00548
constant	-5.955	-6.465	-7.171	-7.299	-7.467	-8.151	-8.523	-9.442

TABLE 4

Minimum k Guaranteeing at Least Specified Accuracy for Modified Peizer Approximation to the Hypergeometric Distribution

Accuracy	.0005	.0001	.00005	.00001	.000005
Any tail probability	4	9	13	29	42
Tail probability ≤.01	2	4	6	16	25
Tail probability ≤.001	2	3	3	8	11

TABLE 5

Maximum Absolute Error of the Modified Peizer Approximation
at k = 8 and Selected Values of (N, n)

N \ n	17	18	19	20	21	22	23	24	25	26	27	28	29	30	32	34	36	38	40	42	44	46	48
35	$^{4}3$[a]	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
36	$^{4}6$	$^{4}6$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
37	$^{4}8$	$^{4}8$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
38	$^{4}9$	$^{4}1$	$^{3}1$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
39	$^{3}1$	$^{3}1$	$^{3}1$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
40	$^{3}1$	$^{3}1$	$^{3}1$	$^{3}1$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
41	$^{3}1$	$^{3}1$	$^{3}1$	$^{3}1$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
42	$^{4}9$	$^{4}9$	$^{4}9$	$^{4}9$	$^{4}9$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
43	$^{4}9$	$^{4}9$	$^{4}9$	$^{4}9$	$^{4}9$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
44	$^{4}8$	$^{4}8$	$^{4}8$	$^{4}9$	$^{4}9$	$^{4}7$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
45	$^{4}8$	$^{4}8$	$^{4}8$	$^{4}8$	$^{4}8$	$^{4}8$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
46	$^{4}8$	$^{4}8$	$^{4}8$	$^{4}8$	$^{4}8$	$^{4}8$	$^{4}5$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
48	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}8$	$^{4}8$	$^{4}8$	$^{4}6$	$^{4}3$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
50	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}5$	$^{4}2$	--	--	--	--	--	--	--	--	--	--	--	--	--	--
52	$^{4}6$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}6$	$^{4}3$	$^{4}1$	--	--	--	--	--	--	--	--	--	--	--	--	--
54	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}4$	$^{4}2$	$^{5}7$	--	--	--	--	--	--	--	--	--	--	--	--
56	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}7$	$^{4}5$	$^{4}3$	$^{4}1$	$^{5}4$	--	--	--	--	--	--	--	--	--	--	--
58	$^{4}5$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}7$	$^{4}6$	$^{4}4$	$^{4}2$	$^{5}8$	$^{5}2$	--	--	--	--	--	--	--	--	--	--
60	$^{4}5$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}5$	$^{4}3$	$^{4}1$	$^{5}4$	$^{5}1$	--	--	--	--	--	--	--	--	--
70	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}5$	$^{4}4$	$^{4}2$	$^{5}3$	$^{6}2$	--	--	--	--	--	--	--
80	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}5$	$^{4}2$	$^{5}5$	$^{6}5$	$^{7}2$	$^{9}3$	--	--	--	--
90	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}5$	$^{4}3$	$^{5}7$	$^{5}1$	$^{7}8$	$^{8}3$	$^{10}4$	--	--
100	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}6$	$^{4}6$	$^{4}6$	$^{4}5$	$^{4}3$	$^{5}9$	$^{5}2$	$^{6}2$	$^{7}1$	$^{9}9$	--
110	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}6$	$^{4}5$	$^{4}3$	$^{5}9$	$^{5}2$	$^{6}3$	$^{7}3$	$^{8}2$
120	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}3$	$^{4}1$	$^{5}2$	$^{6}4$	$^{7}5$
130	$^{4}3$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}4$	$^{4}2$	$^{5}9$	$^{5}3$	$^{6}5$
140	$^{4}3$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}4$	$^{4}2$	$^{5}9$	$^{5}3$
150	$^{4}3$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}4$	$^{4}2$	$^{5}8$
160	$^{4}3$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}3$	$^{4}2$
180	$^{4}3$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}4$
200	$^{4}3$	$^{4}3$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$
220	$^{4}3$	$^{4}3$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$
240	$^{4}3$	$^{4}3$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}4$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}5$

TABLE 5 (contd.)

Maximum Absolute Error of the Modified Peizer Approximation at k = 8 and Selected Values of (N, n)

N \ n	50	52	54	56	58	60	62	64	66	68	70	72	74	76	78	80
110	$^{10}4$	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
120	$^{8}4$	$^{9}2$	--	--	--	--	--	--	--	--	--	--	--	--	--	--
130	$^{7}8$	$^{8}8$	$^{9}5$	$^{10}2$	--	--	--	--	--	--	--	--	--	--	--	--
140	$^{6}6$	$^{6}1$	$^{7}1$	$^{8}1$	$^{10}6$	--	--	--	--	--	--	--	--	--	--	--
150	$^{5}2$	$^{6}6$	$^{6}1$	$^{7}2$	$^{8}2$	$^{9}1$	--	--	--	--	--	--	--	--	--	--
160	$^{5}7$	$^{5}2$	$^{6}6$	$^{6}1$	$^{7}2$	$^{8}2$	$^{9}2$	--	--	--	--	--	--	--	--	--
180	$^{4}2$	$^{4}1$	$^{5}5$	$^{5}2$	$^{6}5$	$^{6}1$	$^{7}2$	$^{8}3$	$^{9}4$	$^{10}4$	--	--	--	--	--	--
200	$^{4}4$	$^{4}3$	$^{4}2$	$^{5}9$	$^{5}4$	$^{5}1$	$^{6}4$	$^{6}1$	$^{7}2$	$^{8}4$	$^{9}5$	$^{10}7$	--	--	--	--
220	$^{4}5$	$^{4}5$	$^{4}3$	$^{4}2$	$^{4}1$	$^{5}6$	$^{5}2$	$^{6}9$	$^{6}3$	$^{7}7$	$^{7}2$	$^{8}3$	$^{9}6$	$^{10}8$	--	--
240	$^{4}5$	$^{4}5$	$^{4}5$	$^{4}4$	$^{4}3$	$^{4}2$	$^{5}8$	$^{5}4$	$^{5}2$	$^{6}6$	$^{6}2$	$^{7}5$	$^{7}1$	$^{8}3$	$^{9}5$	$^{10}8$

a $^{c}d = .0^{c}d = .d\times10^{-c}$

TABLE 6

Maximum N Searched Exhaustively ($2k+1 \le n \le r \le N-n$) and in Region of Restricted Search ($r-n \le 12$)

k	0	1	2	3	4	5	6	7	
Exhaustive	400	400	400	400	400	400	400	375	
k	8	9	10	11	12	13	14	15	
Exhaustive	450	450	500	450	500	450	500	450	
k	16	18	20	24	28	32	36	40	50
Exhaustive	475	500	500	550	650	400	400	750	650
Restricted	500	550	650	750	850	1000	1100	1200	1500

TABLE 7

Location of Maximum Absolute Error of Modified Peizer Approximation to the Hypergeometric Distribution at k = 8 and k = 16

min(P,1-P)	≤ .5 > .1	.10 .05	.05 .01	.010 .005	.005 .001	.0010 .0005	.0005 .0001	.0001 0
k = 8								
n at max	20	19	20	29	27	24	23	21
r at max	20	24	26	31	28	26	25	22
N at max	40	43	46	218	199	199	198	187
k = 16								
n at max	37	36	39	62	60	54	52	50
r at max	37	42	42	64	62	54	55	53
N at max	74	78	81	396	394	347	359	373

Figure A. Maximum Absolute Errors of Approximations

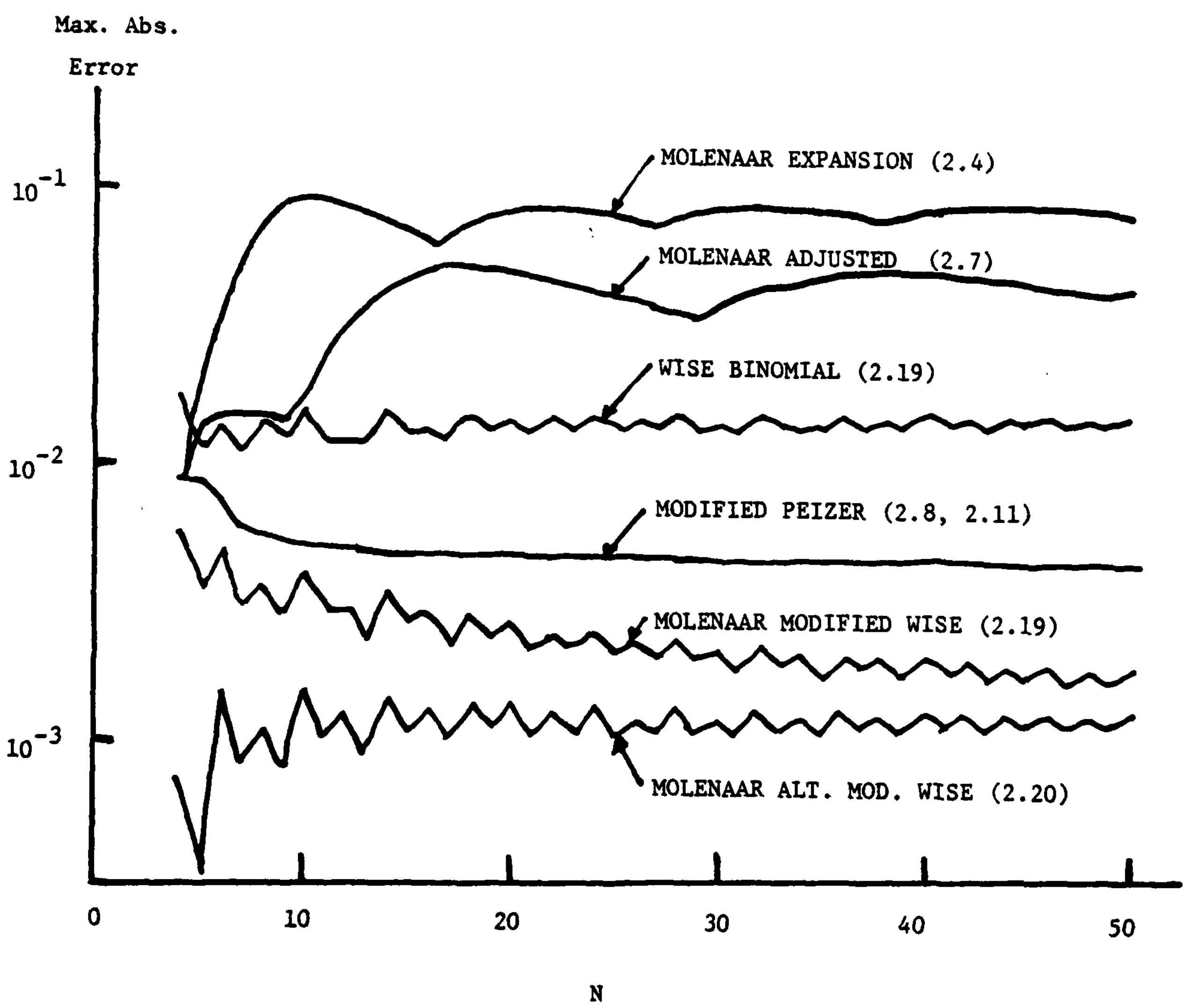


Figure B.

Maximum Absolute Error of Modified Peizer
Approximation to the Hypergeometric Distribution

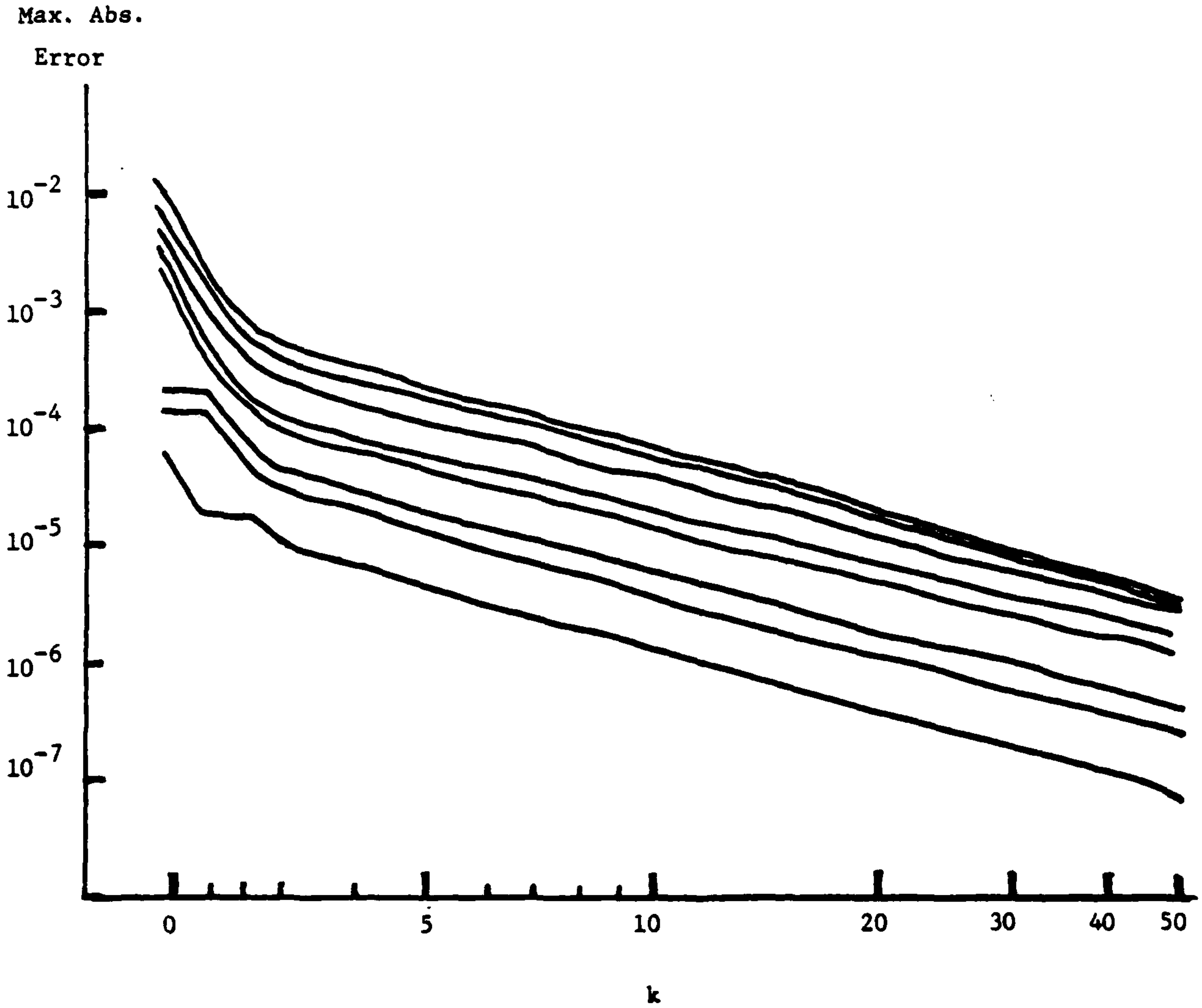


Figure C.

Contours of Maximum Absolute Error of
the Modified Peizer Approximation for k = 8

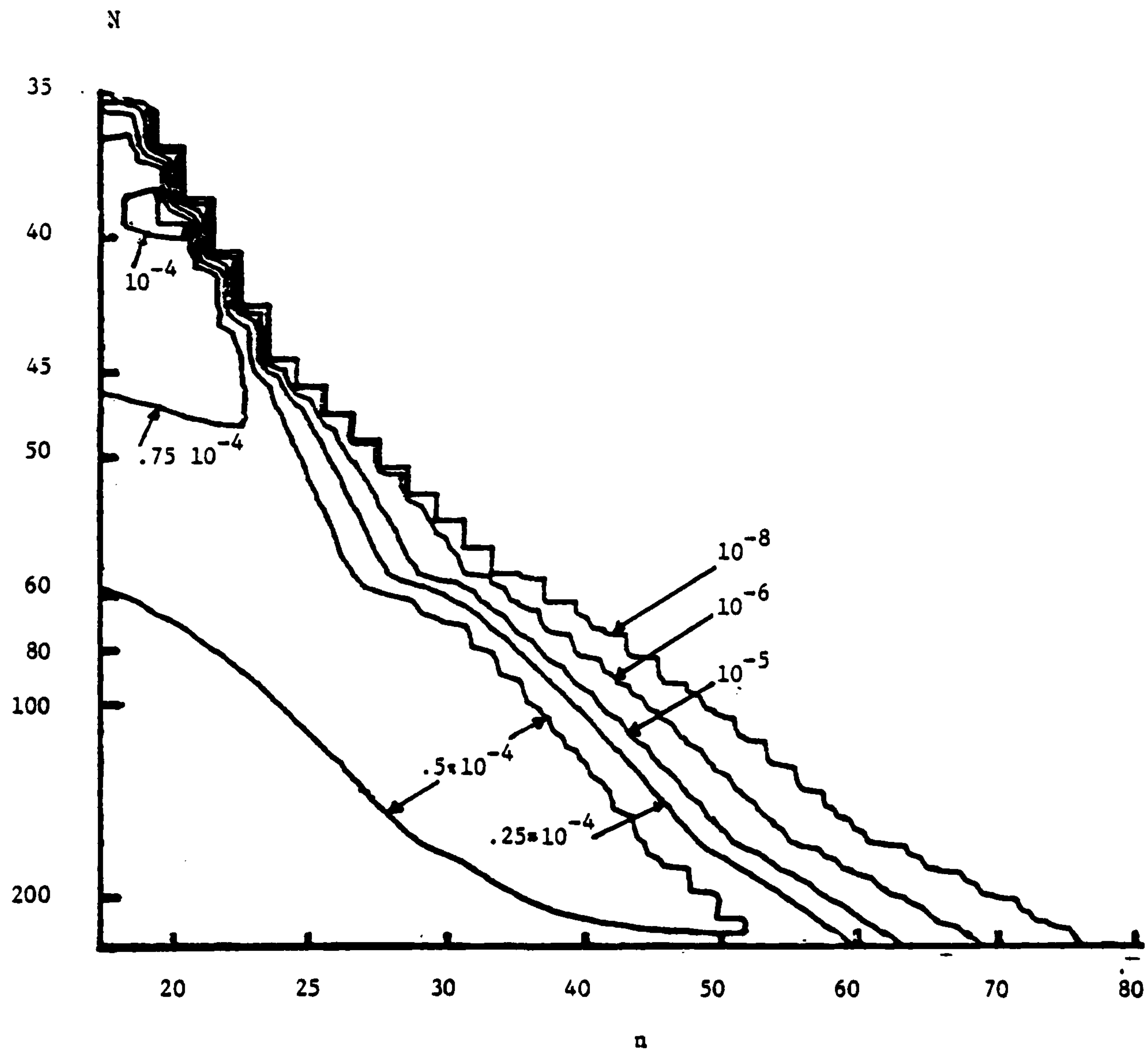
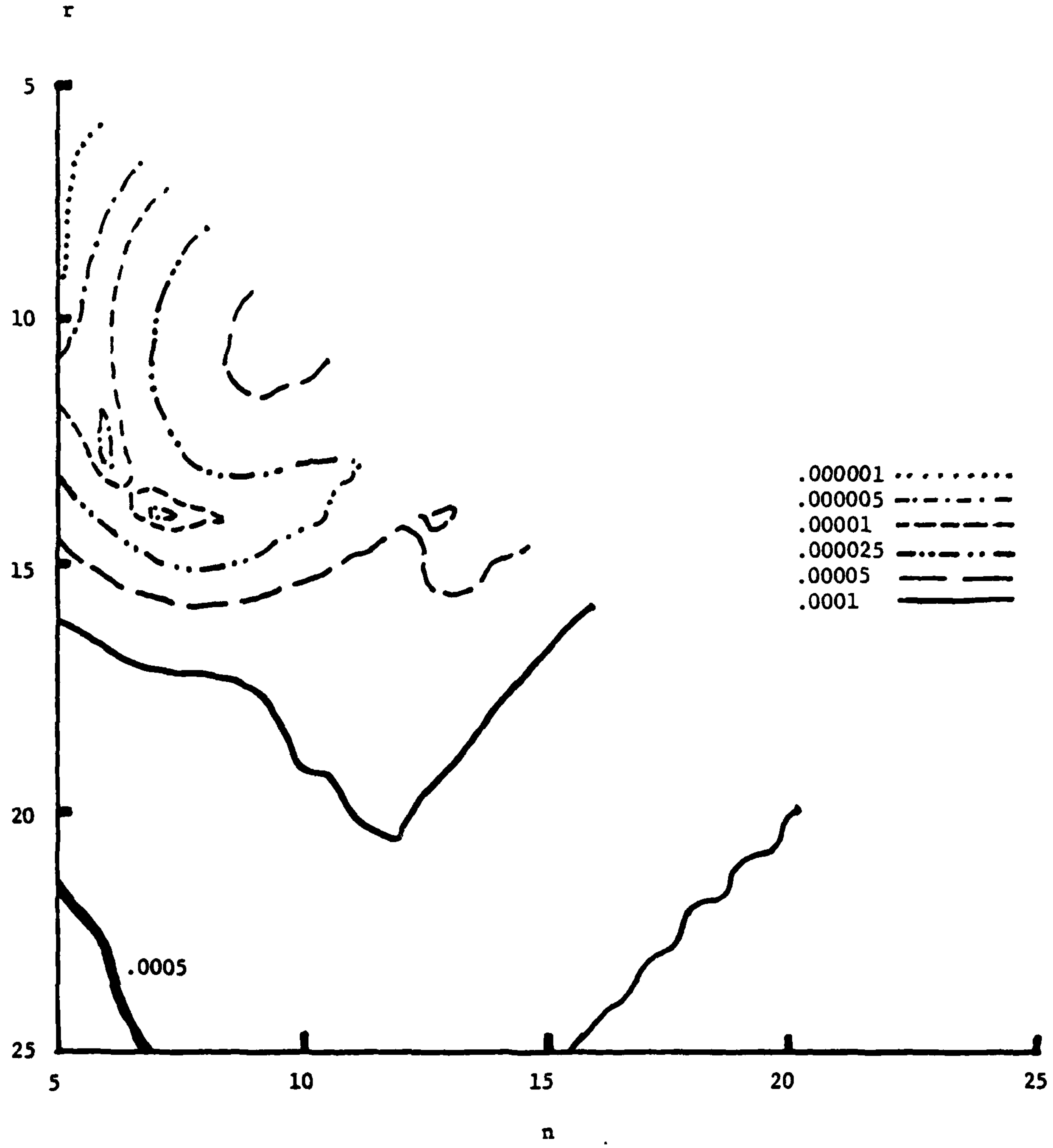


Figure D.

Contours of Maximum Absolute Error of the Modified Peizer Approximation for $a \geq 1$, $N = 50$, and $a \leq n \leq r \leq 25$



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ~~N115~~ N126	2. GOVT ACCESSION NO. AD-A097541	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Accuracy of a Modified Peizer Approximation to the Hypergeometric Distribution, With Comparisons to Some Other Approximations		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER TR #348
7. AUTHOR(s) Robert F. Ling - Clemson University; John W. Pratt - Harvard University		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0451
9. PERFORMING ORGANIZATION NAME AND ADDRESS Clemson University, Dept. of Mathematical Sciences, Clemson, South Carolina 29631		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 047-202
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Code 434, Arlington, Va. 22217		12. REPORT DATE July, 1980
		13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Maximum absolute error, Hypergeometric distribution, Normal approximation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Results of an extensive empirical study of the accuracy of seven normal and three binomial approximations to the hypergeometric distribution are presented in terms of maximum absolute error under various conditions on the variables. The most useful condition are provided by the minimum cell in the given or complementary 2 2 table and the tail probability itself. Of the normal approximations, a modification on one due to Peizer is far the best. It has error at most .0001, for example, if the minimum cell is at least 9, or if the tail probability is below .01 and the minimum cell is at >

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

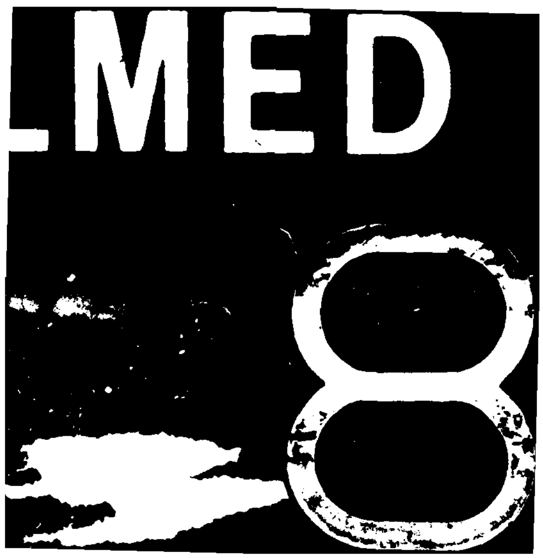
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

#20 (continued).

least 4. Especially detailed results are given for this approximation.


LMED
8